

CORRELATION BOUND FOR DISTANT PARTS OF FACTOR OF IID PROCESSES

ÁGNES BACKHAUSZ, BALÁZS GERENCSÉR, VIKTOR HARANGI, AND MÁTÉ VIZER

ABSTRACT. We study factor of i.i.d. processes on the d -regular tree for $d \geq 3$. We show that if such a process is restricted to two distant connected subgraphs of the tree, then the two parts are basically uncorrelated. More precisely, any functions of the two parts have correlation at most $k(d-1)/(\sqrt{d-1})^k$, where k denotes the distance of the subgraphs. This result can be considered as a quantitative version of the fact that factor of i.i.d. processes have trivial 1-ended tails.

1. INTRODUCTION

This paper deals with factor of i.i.d. processes on the d -regular tree T_d for $d \geq 3$. Loosely speaking, we first put independent and identically distributed (say $[0, 1]$ uniform) random labels on the vertices of T_d ; then each vertex gets a new label that depends on the labelled rooted graph as seen from that vertex, all vertices “using the same rule”.

For a formal definition, let $V(T_d)$ denote the vertex set and $\text{Aut}(T_d)$ the automorphism group of T_d . Suppose that M is a measurable space and $F: [0, 1]^{V(T_d)} \rightarrow M^{V(T_d)}$ is a measurable function. Then F is said to be an $\text{Aut}(T_d)$ -factor (or factor in short) if it is $\text{Aut}(T_d)$ -equivariant, that is, it commutes with the natural $\text{Aut}(T_d)$ -actions. (In most applications M is either a discrete set or \mathbb{R} .) Let $\pi: M^{V(T_d)} \rightarrow M$ denote the coordinate projection corresponding to a distinguished vertex. Then the function $f = \pi \circ F: [0, 1]^{V(T_d)} \rightarrow M$ (often called the *rule*) will be invariant under the stabilizer of the distinguished vertex. It is easy to see that F is determined by f .

If we have an i.i.d. process $Z = (Z_v)_{v \in V(T_d)}$ on $[0, 1]^{V(T_d)}$, then applying F yields a factor of i.i.d. process $X = F(Z)$, which can be viewed as a collection $X = (X_v)_{v \in V(T_d)}$ of M -valued random variables. It follows immediately from the definition that the distribution of X is invariant under the action of $\text{Aut}(T_d)$; in particular, each X_v has the same distribution.

A natural question is “how independent” the random variables X_v are. It is fairly easy to see that for any factor the correlation of X_u and X_v converges to 0 as the distance of u and v goes to infinity. In [4] the following sharp bound was found for the correlation:

$$(1) \quad |\text{corr}(X_u, X_v)| \leq \left(k + 1 - \frac{2k}{d}\right) \left(\frac{1}{\sqrt{d-1}}\right)^k, \text{ where } k = \text{dist}(u, v),$$

2010 *Mathematics Subject Classification.* 60K35, 37A50.

Key words and phrases. Factor of IID, invariant processes, regular tree, correlation, tail sigma-algebra, non-backtracking operator.

The first author was partially supported by the Hungarian Scientific Research Fund (OTKA, grant no. K109684), and the MTA Rényi Institute “Lendület” Limits of Structures Research Group. The third author was supported by Marie Skłodowska-Curie Individual Fellowship Grant No. 661025 and the MTA Rényi Institute “Lendület” Groups and Graphs Research Group. The fourth author was supported by Hungarian National Scientific Fund, grant number: SNN-116095, and was partially supported by ERC Consolidator Grant 648017.

that is, the rate of the correlation decay is essentially $1/(\sqrt{d-1})^k$. (Here it is assumed that $M = \mathbb{R}$ and $\text{var } X_v < \infty$.)

1.1. Results. The main result of this paper basically says that if two connected subsets $V_1, V_2 \subset V(T_d)$ have large distance, then they are “almost independent” in the following sense: for an arbitrary factor X , any function of $(X_v)_{v \in V_1}$ and any function of $(X_v)_{v \in V_2}$ have small correlation, essentially of (the optimal) order $1/(\sqrt{d-1})^k$.

Theorem 1.1. *Let $X = (X_v)_{v \in V(T_d)}$ be a factor of i.i.d. process on $M^{V(T_d)}$ for some measurable space M , and let $V_1, V_2 \subset V(T_d)$ be arbitrary (possibly infinite) subsets of the vertex set. Suppose that $h_i: M^{V_i} \rightarrow \mathbb{R}$ is a measurable function, $i = 1, 2$. For $h_i((X_v)_{v \in V_i})$ we simply write $h_i(X)$. If $h_1(X)$ and $h_2(X)$ have finite variances, then we have*

$$(2) \quad |\text{corr}(h_1(X), h_2(X))| \leq k(d-1) \left(\frac{1}{\sqrt{d-1}} \right)^k,$$

where k denotes the distance of the convex hulls of V_1 and V_2 , which we assume to be positive. (The convex hull of V_i is the smallest connected subgraph of T_d containing V_i .)

One might wonder if a similar bound could exist if k denoted the distance of V_1 and V_2 instead of the distance of the convex hulls. The simplest case where this would make a difference is when V_1 consists of the two endpoints of a path of length $2k$ and V_2 is the one-element set containing the midpoint of this path. Then the distance of V_1 and V_2 is k , while the distance of the convex hulls is 0. In this case the above theorem would be of no use. Can we still have a good bound for the correlation? The answer is negative, as the correlation might actually be 1 in this case for any k . This will be shown by Example 2.3, where we will construct a factor of i.i.d. process X on $[0, 1]^{V(T_d)}$ with the property that X_u and X_v determine the values of X along the whole path connecting u and v . In fact, this process will show that (2) is essentially sharp: for any V_1, V_2 there exist h_1, h_2 such that the correlation in question is of order $1/(\sqrt{d-1})^k$ where k is the distance of the convex hulls.

Given an invariant process $(X_v)_{v \in V(T_d)}$, an event is in the *tail* of the process if it is, for each $r \in \mathbb{N}$, contained by the σ -algebra generated by the random variables X_v for vertices v outside the r -ball around a fixed root. It is open whether any invariant process with a trivial tail can be obtained as some factor of an i.i.d. process. As for the other direction, there are examples for factor of i.i.d. processes whose tail is not trivial. In fact, they can even have full tails. (See [24, Proposition 2.4] or Example 2.3 of the current paper.) Our result is related to another kind of tails called the *1-ended tails*: given an infinite path starting at the root, consider those events that are, for each $r \in \mathbb{N}$, contained by the σ -algebra generated by X_v 's for vertices v of T_d that are separated from the root by the r th vertex of the path.

Let us consider Theorem 1.1 in the special case when h_1, h_2 are indicator functions of two events with one event being fixed and the other running through a sequence of events in such a way that the distance k goes to infinity. Then we obtain that the sequence is “asymptotically independent” from the fixed event. This is actually equivalent to the triviality of the 1-ended tails, see Remark 2.6. Therefore the following is an immediate consequence of Theorem 1.1.

Corollary 1.2. *The 1-ended tail σ -algebras are trivial for any factor of i.i.d. process on $M^{V(T_d)}$ for $d \geq 3$.*

As pointed out by Russell Lyons [24, Section 2], this corollary was known more generally: Pemantle showed that basically every ergodic invariant process on T_d has trivial 1-ended tails [27, Theorem 1]. See [27, Corollary 7] for an equivalent formulation (using asymptotic independence as described above). Our main theorem provides a quantitative version of this formulation in the case of factor of i.i.d. processes: we obtain a universal bound for an arbitrary factor and for arbitrary events (only depending on the distance k). It is also worth mentioning that the triviality of 1-ended tails implies mixing and a weak law of large numbers [27, Corollary 8-10].

Corollary 1.2 will be complemented by Example 2.3, where we construct a factor of i.i.d. process for which, loosely speaking, “any tail broader than a 1-ended tail” is non-trivial. This means that Corollary 1.2 is, in some sense, the best we can hope for.

Note that Corollary 1.2 is also true for $d = 2$ (that is, on \mathbb{Z}) by a result of Rokhlin and Sinai [31].

We finish this section with a brief outline of the proof of Theorem 1.1. Let \tilde{T}_{d-1} denote the rooted $(d-1)$ -ary tree. It is easy to see that we might assume that V_1 and V_2 are both isomorphic to \tilde{T}_{d-1} and their roots have distance k . Let e_i be the directed edge starting at the root of V_i and “pointing away” from V_i . We will show that $|\text{corr}(h_1(X), h_2(X))|$ is maximized by functions h_1 and h_2 that are invariant under the automorphism group of \tilde{T}_{d-1} . In fact, they should “come from” the same $\text{Aut}(\tilde{T}_{d-1})$ -invariant measurable function $f: M^{V(\tilde{T}_{d-1})} \rightarrow \mathbb{R}$. Given such a function f , an $\text{Aut}(T_d)$ -invariant process on the vertices of T_d can be turned into an $\text{Aut}(T_d)$ -invariant process on the directed edges of T_d : for any directed edge e of T_d , apply f to the (labelled) subgraph “behind” e and write its value on e . The process we obtain on the directed edge set $E(T_d)$ will be a factor of i.i.d. process (see Section 2 for precise definitions) whose value on e_i is $h_i(X)$, $i = 1, 2$. Therefore to complete the proof we need to prove a correlation decay result similar to (1) but for directed edges instead of vertices (see Theorem 4.1). To this end we will need to estimate the norms of the powers of the non-backtracking operator (see Section 4.1).

1.2. Related work. Factor of i.i.d. processes can be viewed from an ergodic theoretic point of view, namely, as factors of the Bernoulli shift. \mathbb{Z} -factors (as part of classical ergodic theory) have the largest literature and the most complete theory. For amenable group actions entropy serves as a complete invariant (for isomorphism of i.i.d. processes). A classical example of Ornstein and Weiss [26] expresses the 4-shift as a factor of the 2-shift over free groups of rank at least 2, showing that no notion of entropy can exist in the non-amenable case that would exhibit all the nice properties of the classical Kolmogorov-Sinai entropy. Nevertheless, various definitions of entropy have been introduced and examined in relation to factor maps, see e.g. [9, 10, 22, 32].

One of the reasons why factor of i.i.d. processes have attracted a growing attention in recent years is that they give rise to some sort of randomized local algorithms that can be carried out on arbitrary regular graphs with “large essential girth”, e.g. random regular graphs. Such factor of i.i.d. constructions include perfect matchings [14, 25], independent sets [12, 13, 18, 19], 4-regular spanning forests [16, 23], colorings [11]. Using this connection to random regular graphs in the reverse direction, one can prove entropy inequalities [3, 9] yielding necessary conditions for a process to be factor of i.i.d.; this new tool has several applications already [17, 29, 30]. Correlation bounds provide further necessary conditions [4]; this is the main tool in the current paper as well. In [5] the possible “correlation structures” were described for factor of i.i.d. processes by understanding their spectral measures. See [24] for further references and for many open problems in the topic.

Outline of the paper. The rest of the paper is structured as follows. In Section 2 we go through basic definitions, present some examples, and show how Corollary 1.2 follows from Theorem 1.1, the proof of which is given in Section 3. Finally, in Section 4 we prove our correlation decay result for directed edges via bounding the norms of the powers of the non-backtracking operator.

2. PRELIMINARIES

2.1. Factors of i.i.d. Suppose that a group Γ acts on a countable set S . Then Γ also acts on the space M^S for a set M : for any function f defined on S and for any $\gamma \in \Gamma$ let

$$(3) \quad (\gamma \cdot f)(s) := f(\gamma^{-1} \cdot s) \quad \forall s \in S.$$

In our setting M is always equipped with a σ -algebra (that is, M is a measurable space). As usual, the product space M^S is equipped with the smallest σ -algebra such that all $M^S \rightarrow M$ coordinate projections are measurable. This way a $\cdot \rightarrow M^S$ function is measurable if and only if each coordinate function is measurable. First we define the notion of factor maps.

Definition 2.1. Let M_1, M_2 be measurable spaces and S_1, S_2 countable sets with a group Γ acting on both. A measurable mapping $F: M_1^{S_1} \rightarrow M_2^{S_2}$ is said to be a Γ -factor if it is Γ -equivariant, that is, it commutes with the Γ -actions.

Whenever Γ acts transitively on S_2 , F is determined by the function $f = \pi_o \circ F: M_1^{S_1} \rightarrow M_2$, where $\pi_o: M_2^{S_2} \rightarrow M_2$ is the projection corresponding to some distinguished element $o \in S_2$. There is a one-to-one correspondence between measurable Γ -equivariant mappings $F: M_1^{S_1} \rightarrow M_2^{S_2}$ and measurable functions $f: M_1^{S_1} \rightarrow M_2$ that are invariant under the stabilizer of o .

Next we explain what we mean by processes on M^S .

Definition 2.2. A probability measure on M^S that is invariant under the Γ -action is called an *invariant process*. The simplest examples are *i.i.d. processes*: take a probability measure μ on M and consider the product measure $\nu = \mu^S$ on M^S . (The Γ -action on (M^S, μ^S) is often called the *generalized Bernoulli shift*.) Given an i.i.d. process ν on $M_1^{S_1}$ and a Γ -factor $F: M_1^{S_1} \rightarrow M_2^{S_2}$, the push-forward measure $F_*\nu$ is also Γ -invariant. Such processes are called *factors* of the i.i.d. process ν .

Sometimes we will think of an invariant process ν on M^S as an M^S -valued random variable (whose distribution is ν), or as a collection of M -valued random variables (whose joint distribution is ν). For example, if Z_s , $s \in S_1$ are independent, M_1 -valued random variables with some common distribution μ , then $Z = (Z_s)_{s \in S_1}$ is an i.i.d. process on $M_1^{S_1}$. Given a Γ -factor $F: M_1^{S_1} \rightarrow M_2^{S_2}$, $X := F(Z)$ is a factor of the i.i.d. process Z . Then X is a collection $(X_s)_{s \in S_2}$ of M_2 -valued random variables. They can be expressed using the corresponding f as well: clearly $X_o = f(Z)$, and it is also easy to see that $X_s = f(\gamma \cdot Z)$, where $\gamma \in \Gamma$ is an arbitrary group element taking o to s .

2.2. Factors on T_d . In this paper we mainly consider the case when Γ is the automorphism group $\text{Aut}(T_d)$ of the d -regular infinite tree T_d ($d \geq 3$) and S is either the vertex set $V(T_d)$ or the directed edge set $E(T_d)$. The latter consists of the ordered pairs (u, v) , where $u, v \in V(T_d)$ are neighbors. For a directed edge $e = (u, v)$, the inverse of e is the directed edge $e^{-1} = (v, u)$.

In this setting, when we say *factor of i.i.d. process*, we do not need to specify which i.i.d. process we have in mind. The reason for this is that i.i.d. processes are factors of each other in this case. The most natural i.i.d. processes to consider would be:

- ν_1 : $S = V(T_d)$; $M = [0, 1]$; μ is the Lebesgue measure,
- ν_2 : $S = V(T_d)$; $M = \{0, 1, \dots, m-1\}$ for some $m \geq 2$; μ is the uniform measure on M ,
- ν_3 : $S = E(T_d)$; $M = [0, 1]$; μ is the Lebesgue measure,
- ν_4 : $S = E(T_d)$; $M = \{0, 1, \dots, m-1\}$ for some $m \geq 2$; μ is the uniform measure on M .

It is trivial that ν_2 is a factor of ν_1 and that ν_4 is a factor of ν_3 . Extending the classical example of Ornstein and Weiss [26], it was shown in [6] that ν_1 is an $\text{Aut}(T_d)$ -factor of ν_2 . (In [6] the author considers the case $m = 2$ but the same argument works for arbitrary $m \geq 2$.) Furthermore, it can be seen easily that ν_2 is a factor of ν_4 . (Let the label of a vertex $v \in V(T_d)$ be the sum of the labels of the directed edges starting at v , modulo m .) Finally, it is possible to obtain ν_3 as a factor of ν_1 . (We can think of the $[0, 1]$ -label of a vertex as $d+1$ independent $[0, 1]$ -labels. For each vertex v , we order its neighbors according to their $(d+1)$ th labels. The directed edge going from v to its i th largest neighbor will get the i th label of v .) These constructions show that the above i.i.d. processes are all factors of each other. We will usually work with ν_1 .

Block factors are $\text{Aut}(T_d)$ -factors obtained using a rule f that depends only on some finite-radius ball around the distinguished vertex $o \in V(T_d)$. If this radius is 0, then we simply have $X_v = \varphi(Z_v), \forall v \in V(T_d)$ for some measurable $\varphi: M_1 \rightarrow M_2$. If this is the case, then we say that X is a *pointwise factor* of Z . If Z is an i.i.d. process, then so is any pointwise factor of Z . However, as the next example shows, there exists a factor of i.i.d. process X that is “universal” in the sense that any factor of i.i.d. process can be obtained as a pointwise factor of X .

Example 2.3. Let Z denote the i.i.d. process on $[0, 1]^{V(T_d)}$ (of distribution ν_1). We claim that there exists a factor X of Z on $[0, 1]^{V(T_d)}$ with the following properties.

- (a) With probability 1 for any $u, v \in V(T_d)$ the values of X_u, X_v determine (in a measurable way) the values of X along the whole path connecting u and v . In other words, for any $V \subset V(T_d)$, $X_v, v \in V$ determine X on the convex hull of V .
- (b) Any factor X' of Z on some $M^{V(T_d)}$ can be obtained as the pointwise factor of X , that is, $X'_v = \varphi(X_v), \forall v \in V(T_d)$ for some measurable $\varphi: [0, 1] \rightarrow M$.

Proof. The idea is to encode the whole labelled tree in X_v . To do this with an $\text{Aut}(T_d)$ -factor we need to do the encoding in a way that it only contains the isomorphism type of the labelled tree rooted at v . With probability 1 the values Z_v are pairwise distinct. Then for a given configuration $\omega = (\omega_u)_{u \in V(T_d)}$ of pairwise disjoint labels we assign the following sequence to any given vertex v :

$$\left(\underbrace{\omega_v}_{\text{label of } v}, \underbrace{\omega_{v_1} < \dots < \omega_{v_d}}_{\text{labels of the neighbors of } v \text{ in increasing order}}, \underbrace{\omega_{v_{1,1}} < \dots < \omega_{v_{1,d-1}}}_{\text{labels of the remaining } d-1 \text{ neighbors of } v_1}, \underbrace{\omega_{v_{2,1}} < \dots < \omega_{v_{2,d-1}}}_{\text{labels of the remaining } d-1 \text{ neighbors of } v_2}, \dots \right).$$

Finally, we apply a fixed injective measurable $[0, 1]^{\mathbb{N}} \rightarrow [0, 1]$ mapping (whose inverse is also measurable) to any such sequence to get the new label α_v of v .

By knowing the labels α_u and α_v of the vertices u and v , we know the isomorphism type of the ω -labelled tree, and also the original labels ω_u, ω_v of u and v . Then to find out

the α -label of any other vertex, it suffices to know its original ω -label. Now let n denote the distance of u and v . Then for any integer $0 < k < n$, the k -ball around u and the $(n - k)$ -ball around v have exactly one common vertex (the k th vertex on the u - v path). Therefore there is one common value among the ω -labels of these two balls, the label of the common vertex. Thus α_u and α_v indeed determine the ω -label and hence the α -label of any vertex on the u - v path, which proves (a).

To see (b), recall that any factor X' of Z comes from a rule $f: [0, 1]^{V(T_d)} \rightarrow M$ that is measurable and invariant under the stabilizer of o . Clearly, such an f can be obtained as a measurable function φ of the α -label at o . \square

We claim that Theorem 1.1 is essentially sharp for any $V_1, V_2 \subset V(T_d)$. Let X be a process satisfying the properties (a) and (b) above, and let us pick v_1 and v_2 in the convex hulls of V_1 and V_2 with the smallest possible distance: $\text{dist}(v_1, v_2) = k$. It follows from (a) that X_{v_i} is a measurable function of $X_u, u \in V_i$, and by (b) we can obtain X'_{v_i} as a measurable function of X_{v_i} for any factor X' of Z . Thus we can choose measurable h_i in a way that $\text{corr}(h_1(X), h_2(X)) = \text{corr}(X'_{v_1}, X'_{v_2})$ for any given factor of i.i.d. process X' on $\mathbb{R}^{V(T_d)}$. Therefore the fact that the vertex-correlation bound (1) is sharp means that (2) is also essentially sharp. (In [4] the authors give a very simple example for which the correlation of two vertices of distance k is of order $1/(\sqrt{d-1})^k$, which already shows that (1) is essentially sharp, but it is clear from their proof of the bound that it actually has to be sharp.)

2.3. Tail σ -algebras. For $v \in V(T_d)$ let $\pi_v: M^{V(T_d)} \rightarrow M$ denote the natural coordinate projection. For $V \subseteq V(T_d)$ let $\sigma(V)$ be the σ -algebra generated by the maps $\pi_v, v \in V$.

Definition 2.4. The *tail σ -algebra* is defined as $\bigcap_r \sigma(V(T_d) \setminus B_r)$, where B_r stands for the r -ball around some fixed vertex o . Clearly, the tail does not depend on the choice of o .

A σ -algebra is said to be *trivial* w.r.t. a probability measure if it contains only sets of measure 0 or 1.

We say that a process *has full tail* if its tail is the whole σ -algebra $\sigma(V(T_d))$.

It is open whether trivial tail implies factor of i.i.d. As for the reverse direction, it follows easily from the Kolmogorov 0-1 Law that block factors have trivial tail. This is not true for arbitrary factors, though. In [24, Proposition 2.4] it was shown that “the uniform random perfect matching on T_d ” has full tail since knowing the matching outside a ball determines it inside as well. (It had been known earlier by Lyons and Nazarov [25] that this process is factor of i.i.d.)

In Example 2.3 we presented a factor of i.i.d. process on $[0, 1]^{V(T_d)}$ for which $\sigma(V)$ coincides with the σ of the convex hull of V for any $V \subset V(T_d)$. Such a process clearly has a full tail. One could consider some other sequence of shrinking subsets of $V(T_d)$ with empty intersection: $V(T_d) \supset V_1 \supset V_2 \supset \dots$ with $\bigcap_n V_n = \emptyset$, and define another notion of tail by considering the σ -algebra $\bigcap_n \sigma(V_n)$. The only case for which such a tail could be trivial for Example 2.3 is when the convex hulls of V_n “converge to infinity”. The only such tails are the 1-ended tails introduced in [27] under the name “one-sided tails”, see also [24, Section 2].

Definition 2.5. The *1-ended tail σ -algebra* corresponding to an infinite simple path (v_0, v_1, v_2, \dots) is $\bigcap_n \sigma(D_n)$, where D_n is the set of vertices closer to v_n than to v_{n-1} .

It is easy to see that for an $\text{Aut}(T_d)$ -invariant measure on $M^{V(T_d)}$ the 1-ended tails are all trivial or none are trivial.

The group $\text{Aut}^+(T_d)$ of “parity-preserving” automorphisms is a subgroup of $\text{Aut}(T_d)$ of index 2. (The vertices of T_d can be partitioned into two classes based on the parity of their distance to a fixed vertex, and an automorphism either takes each class into itself, i.e. *preserves parity*, or swaps the two classes.) In [27] Pemantle showed that any $\text{Aut}^+(T_d)$ -invariant process that is ergodic has trivial 1-ended tails, which in turn implies mixing, a weak law of large numbers, and a version of the Birkhoff Ergodic Theorem on T_d .

Next we show how the 1-ended tail triviality (for the case of factor of i.i.d. processes) follows from Theorem 1.1.

Proof of Corollary 1.2. Let A_1 be an arbitrary event in the 1-ended tail and $A_2 \in \sigma(V_2)$ for some finite set $V_2 \subset V(T_d)$. We claim that A_1 and A_2 are independent. To see this, it suffices to show that the indicator functions $h_i := \mathbb{1}_{A_i}$, $i = 1, 2$ have zero correlation. However, for any fixed n we can apply Theorem 1.1 for $V_1 = D_n$ since $h_1 = \mathbb{1}_{A_1}$ can be considered as a measurable $M^{D_n} \rightarrow \mathbb{R}$ function for any n . As $n \rightarrow \infty$ the distance of D_n and the convex hull of V_2 gets arbitrarily large, so we obtain that the correlation must be 0.

Therefore A_1 is independent from the σ -algebra $\sigma(V_2)$ for any finite subset V_2 of the vertex set. The Dynkin π - λ lemma implies that A_1 is independent from the whole σ -algebra $\sigma(V(T_d))$. In particular, A_1 is independent from itself, so its probability is either 0 or 1. This holds for any event in the 1-ended tail meaning that it is trivial. \square

Remark 2.6. The following are equivalent for any $\text{Aut}(T_d)$ -invariant process μ on $M^{V(T_d)}$.

- (a) The 1-ended tails are trivial w.r.t. μ .
- (b) Let A denote an event on $M^{V(T_d)}$ depending only on the coordinates $U \subset V(T_d)$ (that is, $A = A' \times M^{V(T_d) \setminus U}$ for some measurable subset A' of M^U). Furthermore, let the events B_n depend on the coordinates V_n . Then B_n is asymptotically independent from A (i.e. $\mu(A \cap B_n) - \mu(A)\mu(B_n) \rightarrow 0$) whenever the distance of the convex hulls of U and V_n goes to infinity as $n \rightarrow \infty$.

The proof above essentially shows that (b) implies (a), while the proof of [27, Corollary 7] yields the other implication.

3. PROOF OF THE MAIN RESULT

Recall that \tilde{T}_{d-1} denotes the rooted $(d-1)$ -ary tree: the degree of each vertex of \tilde{T}_{d-1} is d except for one vertex (the root) of degree $d-1$ (that is, every vertex has $d-1$ “offsprings”). If we delete an edge of T_d , both connected components will be isomorphic to \tilde{T}_{d-1} .

Step 1. Let V_1 and V_2 be subsets of $V(T_d)$ and let k denote the distance of their convex hulls as in Theorem 1.1. This means that there exist unique vertices v_1, v_2 such that v_i is in the convex hull of V_i and there is a (unique) path of length k connecting v_1 and v_2 (we assume $k \geq 1$). If we delete the edges of this path, then V_i is contained in the connected component of v_i , which is isomorphic to \tilde{T}_{d-1} , $i = 1, 2$, see Figure 1. Let us replace V_i with its component. This way V_1 and V_2 get larger while k remains the same, therefore it suffices to prove Theorem 1.1 in this case.

Step 2. So from this point on we will assume that V_1 and V_2 are disjoint copies of \tilde{T}_{d-1} in T_d , with their roots at distance k . This means that $(X_v)_{v \in V_1}$ and $(X_v)_{v \in V_2}$ can be viewed as processes on $V(\tilde{T}_{d-1})$. More precisely, we define $M^{V(\tilde{T}_{d-1})}$ -valued random variables X_1, X_2 by fixing graph isomorphisms $\Phi_i: V(\tilde{T}_{d-1}) \rightarrow V_i$ for $i = 1, 2$, and setting $(X_i)_u := X_{\Phi_i(u)}$ for $i = 1, 2$, $u \in V(\tilde{T}_{d-1})$. Using the $\text{Aut}(T_d)$ -invariance of X the following properties of X_1 and X_2 follow easily.

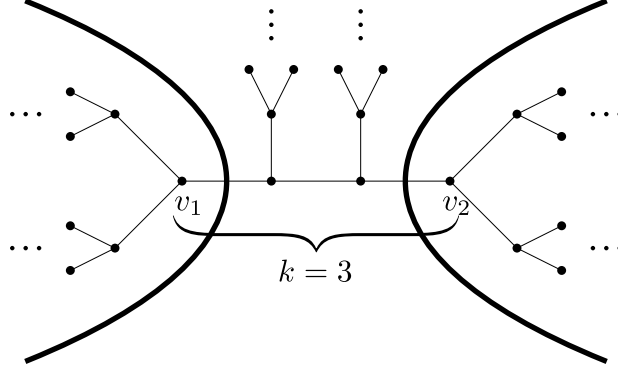


FIGURE 1. Step 1: we might assume that V_1 and V_2 are $(d-1)$ -ary subtrees of distance k

Claim 3.1. *Both X_1 and X_2 are invariant under $\tilde{\Gamma} := \text{Aut}(\tilde{T}_{d-1})$. In fact, $(\gamma_1 \cdot X_1, \gamma_2 \cdot X_2)$ and (X_1, X_2) have the same joint distribution for any $\gamma_1, \gamma_2 \in \tilde{\Gamma}$. Furthermore, (X_1, X_2) and (X_2, X_1) also have the same joint distribution.*

Now let h_1, h_2 be as in Theorem 1.1. Then there clearly exist measurable functions $f_1, f_2: M^{V(\tilde{T}_{d-1})} \rightarrow \mathbb{R}$ such that $f_i(X_i) = h_i(X)$ ($i = 1, 2$). According to the following lemma we might assume that f_1 and f_2 are actually the same function.

Lemma 3.2. *Let (A, \mathcal{F}) be an arbitrary measurable space. Suppose that the (A, \mathcal{F}) -valued random variables X_1, X_2 are exchangeable (that is, (X_1, X_2) and (X_2, X_1) have the same joint distribution), and that there exists a constant $\alpha \geq 0$ with the property that for any measurable $f: A \rightarrow \mathbb{R}$ we have*

$$(4) \quad |\text{corr}(f(X_1), f(X_2))| \leq \alpha \text{ provided that } f(X_1) \text{ has finite variance.}$$

Then for any measurable functions $f_1, f_2: A \rightarrow \mathbb{R}$

$$(5) \quad |\text{corr}(f_1(X_1), f_2(X_2))| \leq \alpha \text{ provided that } f_1(X_1) \text{ and } f_2(X_2) \text{ have finite variances.}$$

Proof. We might assume that $\text{var}(f_1(X_1)) = \text{var}(f_2(X_2)) = 1$. (If one of the variances is 0, then the correlation is 0 by definition and the statement of the lemma holds trivially. Otherwise we can rescale f_1 and f_2 to make the variances equal to 1 without changing the correlation.)

Since X_1 and X_2 are exchangeable we have $\text{cov}(f_1(X_1), f_2(X_2)) = \text{cov}(f_1(X_2), f_2(X_1))$. It follows that

$$\begin{aligned} \text{corr}(f_1(X_1), f_2(X_2)) &= \text{cov}(f_1(X_1), f_2(X_2)) \\ &= \frac{1}{4} \left(\text{cov}((f_1 + f_2)(X_1), (f_1 + f_2)(X_2)) - \text{cov}((f_1 - f_2)(X_1), (f_1 - f_2)(X_2)) \right). \end{aligned}$$

Using the triangle inequality and applying (4) to the function $f = f_1 + f_2$ and to $f = f_1 - f_2$ we obtain that

$$\begin{aligned} |\text{corr}(f_1(X_1), f_2(X_2))| &\leq \frac{\alpha}{4} \left(\text{var}((f_1 + f_2)(X_1)) + \text{var}((f_1 - f_2)(X_1)) \right) \\ &= \frac{\alpha}{4} \left(2 \text{var}(f_1(X_1)) + 2 \text{var}(f_2(X_1)) \right) = \alpha. \end{aligned}$$

□

Step 3. It remains to bound $\text{corr}(f(X_1), f(X_2))$ for any given measurable function $f: M^{V(\tilde{T}_{d-1})} \rightarrow \mathbb{R}$. We claim that it suffices to do this in the case when f is $\tilde{\Gamma} = \text{Aut}(\tilde{T}_{d-1})$ -invariant. The idea is to “average f over the orbits of the $\tilde{\Gamma}$ -action”, that is, take the function $\bar{f}(\omega) = \int_{\tilde{\Gamma}} f(\gamma \cdot \omega) d\gamma$ instead.

To make this more precise, let us consider the natural topology and the *Haar measure* on $\tilde{\Gamma}$. (For $u, v \in V(\tilde{T}_{d-1})$ let $\tilde{\Gamma}_{u,v}$ denote the set of those $\gamma \in \tilde{\Gamma}$ that take u to v . Then the sets $\tilde{\Gamma}_{u,v}$ form a base of the topology. Furthermore, $\tilde{\Gamma}_{u,v}$ is non-empty if and only if u and v have the same distance n from the root of \tilde{T}_{d-1} , in which case the measure of $\tilde{\Gamma}_{u,v}$ is $1/(d-1)^n$.) By μ we will denote the common distribution of X_1 and X_2 , so μ is a probability measure on $M^{V(\tilde{T}_{d-1})}$. It can be seen easily that the following $\tilde{\Gamma} \times M^{V(\tilde{T}_{d-1})} \rightarrow \mathbb{R}$ function is measurable:

$$\varphi(\gamma, \omega) := f(\gamma \cdot \omega) \quad \gamma \in \tilde{\Gamma}, \omega \in M^{V(\tilde{T}_{d-1})},$$

and that $|\varphi|$ has finite integral. Therefore

$$\bar{f}(\omega) := \int_{\tilde{\Gamma}} \varphi(\gamma, \omega) d\gamma = \int_{\tilde{\Gamma}} f(\gamma \cdot \omega) d\gamma$$

is defined for μ -a.e. $\omega \in M^{V(\tilde{T}_{d-1})}$ and it is measurable. Furthermore, $\bar{f}: M^{V(\tilde{T}_{d-1})} \rightarrow \mathbb{R}$ is clearly $\tilde{\Gamma}$ -invariant. (To be more precise, there exists a measurable, $\tilde{\Gamma}$ -invariant $M^{V(\tilde{T}_{d-1})} \rightarrow \mathbb{R}$ function that is equal to \bar{f} μ -a.e.)

Although we will not need this, we mention that there is another way to define \bar{f} : take the σ -algebra of $\tilde{\Gamma}$ -invariant measurable sets in $M^{V(\tilde{T}_{d-1})}$ and let \bar{f} be the conditional expectation of f w.r.t. this σ -algebra.

Proposition 3.3. *The function \bar{f} has the following properties.*

- (a) $\mathbb{E}\bar{f}(X_i) = \mathbb{E}f(X_i)$,
- (b) $\mathbb{E}(\bar{f})^2(X_i) \leq \mathbb{E}f^2(X_i)$,
- (c) $\mathbb{E}\bar{f}(X_1)\bar{f}(X_2) = \mathbb{E}f(X_1)f(X_2)$.

Proof. If ω is μ -random element of $M^{V(\tilde{T}_{d-1})}$, then the distribution of $\gamma \cdot \omega$ is also μ for any fixed $\gamma \in \tilde{\Gamma}$. Using this fact and Fubini's theorem (a) easily follows:

$$\mathbb{E}\bar{f}(X_i) = \int \bar{f} d\mu = \int \int_{\tilde{\Gamma}} f(\gamma \cdot \omega) d\gamma d\mu = \int_{\tilde{\Gamma}} \underbrace{\int f(\gamma \cdot \omega) d\mu}_{\mathbb{E}f(X_i)} d\gamma = \mathbb{E}f(X_i).$$

To see (b) we need to first use the Cauchy-Schwarz inequality before applying Fubini's theorem to φ^2 :

$$\begin{aligned} \mathbb{E}(\bar{f})^2(X_i) &= \int (\bar{f})^2 d\mu = \int \left(\int_{\tilde{\Gamma}} f(\gamma \cdot \omega) d\gamma \right)^2 d\mu \leq \int \int_{\tilde{\Gamma}} f^2(\gamma \cdot \omega) d\gamma d\mu \\ &= \int_{\tilde{\Gamma}} \underbrace{\int f^2(\gamma \cdot \omega) d\mu}_{\mathbb{E}f^2(X_i)} d\gamma = \mathbb{E}f^2(X_i). \end{aligned}$$

Finally, to prove (c) we use that $(\gamma_1 \cdot X_1, \gamma_2 \cdot X_2)$ and (X_1, X_2) have the same joint distribution for any fixed $\gamma_1, \gamma_2 \in \tilde{\Gamma}$. It follows that

$$\mathbb{E}f(\gamma_1 \cdot X_1)f(\gamma_2 \cdot X_2) = \mathbb{E}f(X_1)f(X_2).$$

Integrating this equality w.r.t. $d\gamma_1 d\gamma_2$ and using Fubini's theorem once again we conclude that $\mathbb{E}\bar{f}(X_1)\bar{f}(X_2) = \mathbb{E}f(X_1)f(X_2)$. (Each time we used it, the conditions of Fubini's theorem were satisfied as $|\varphi|$ and φ^2 are measurable and have finite integrals.) \square

It follows from (a) and (c) that $\bar{f}(X_1)$ and $\bar{f}(X_2)$ have the same covariance as $f(X_1)$ and $f(X_2)$, while (a) and (b) imply that $\text{var } \bar{f}(X_i) \leq \text{var } f(X_i)$. Consequently,

$$|\text{corr}(f(X_1), f(X_2))| \leq |\text{corr}(\bar{f}(X_1), \bar{f}(X_2))|.$$

Therefore it suffices to bound the correlation for \bar{f} , that is, we might assume that f was $\tilde{\Gamma}$ -invariant in the first place.

Step 4. Whenever we have a factor of i.i.d. process X_v on $M^{V(T_d)}$ and a measurable $\tilde{\Gamma}$ -invariant function $f: M^{V(\tilde{T}_{d-1})} \rightarrow \mathbb{R}$, we can combine them to create a factor of i.i.d. process Y_e on $\mathbb{R}^{E(T_d)}$. To get Y_e for a directed edge $e = (u, w)$ we take the subtree T_e “behind” e (that is, $V(T_e)$ consists of those vertices of T_d that are closer to u than to w), and apply f to $(X_v)_{v \in V(T_e)}$, see Figure 2. We can do this because T_e is isomorphic to \tilde{T}_{d-1} , and Y_e will be well defined since f is $\tilde{\Gamma}$ -invariant. It is also easy to see that $(Y_e)_{e \in E(T_d)}$ will be a factor of i.i.d. process.

Furthermore, $Y_{e_i} = f(X_i)$, $i = 1, 2$, where e_1 is the directed edge starting at v_1 and “pointing towards” v_2 , and e_2 is the directed edge starting at v_2 and “pointing towards” v_1 . So it remains to show that the correlation of Y_{e_1} and Y_{e_2} is small if the distance of e_1 and e_2 is large. This final step will be done in the next section, see Theorem 4.1 below. The bound (6) clearly implies Theorem 1.1. (Note that $\text{dist}(e_1, e_2) = k - 1$ in our case.)

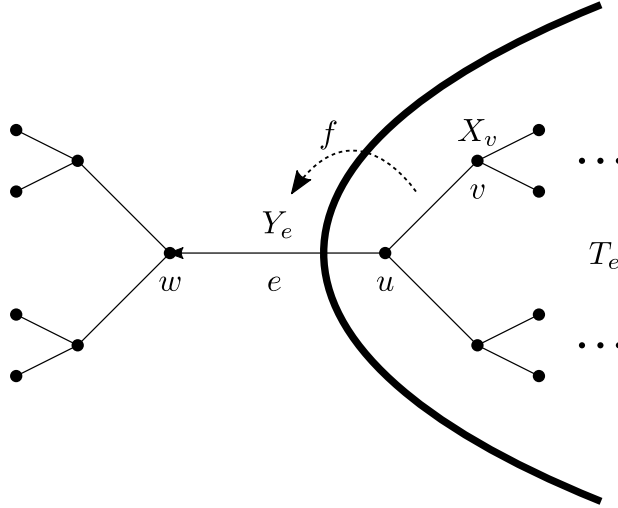


FIGURE 2. Step 4: obtaining Y_e from $(X_v)_{v \in V(T_e)}$

4. CORRELATION DECAY FOR DIRECTED EDGES

In [4] Backhausz, Szegedy and Virág bounded the correlation of a pair of vertices for factor of i.i.d. processes on $\mathbb{R}^{V(T_d)}$, see (1). The goal of this section is to prove a similar bound but for directed edges instead of vertices, that is, for factor of i.i.d. processes on $\mathbb{R}^{E(T_d)}$.

By the distance of two undirected edges e_1, e_2 we mean the smallest integer k for which there exists a path containing $k + 1$ edges including e_1 and e_2 , see Figure 3. As for the

distance of directed edges, we simply forget the directions of the edges and take the distance of the corresponding undirected edges. Equivalently,

$$\text{dist}((u_1, u_2); (v_1, v_2)) := \begin{cases} 0, & \text{if } u_1 = v_1; u_2 = v_2 \text{ or } u_1 = v_2; u_2 = v_1, \\ 1 + \min_{i,j \in \{1,2\}} \text{dist}(u_i, v_j), & \text{otherwise.} \end{cases}$$

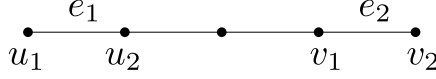


FIGURE 3. The distance of the edges e_1 and e_2 in the figure is 3

Theorem 4.1. *Let $Y = (Y_e)_{e \in E(T_d)}$ be a factor of i.i.d. process on $\mathbb{R}^{E(T_d)}$. Then*

$$(6) \quad |\text{corr}(Y_{e_1}, Y_{e_2})| \leq (k+1) \left(\frac{1}{\sqrt{d-1}} \right)^{k-1}, \text{ where } k = \text{dist}(e_1, e_2),$$

provided that $\text{var } Y_e < \infty$.

It would be possible to start with the vertex-correlation bound (1) and deduce a somewhat weaker version of (6) from that. This would involve some tedious calculations, however. Instead, we will apply similar ideas as in [4], where the norms of certain polynomials of the adjacency operator were determined to obtain the bound (1). Here we will need to work with the non-backtracking operator instead of the adjacency operator.

The key observation is that $\mathbb{E}Y_{e_1}Y_{e_2}$ can be expressed as an inner product on the L^2 space over $[0, 1]^{V(T_d)}$. Some kind of a non-backtracking operator can be defined on this space, and to bound the inner product in question one needs to determine the norm of the k -th power of this operator. This, however, can be traced back to the case of the ordinary non-backtracking operator B on T_d .

4.1. The non-backtracking operator on T_d . For an undirected simple graph G let $V(G)$ and $E(G)$ be the vertex set and the directed edge set of G , respectively. (We assume that G is locally finite.) For $e, e' \in E(G)$ we write $e \rightarrow e'$ if $e = (u, v)$ and $e' = (v, w)$ for some $u, v, w \in V(G)$ with $u \neq w$, that is, if the head of e coincides with the tail of e' and $e' \neq e^{-1}$. If this is the case, then we say that e is the *predecessor* of e' , and e' is the *successor* of e , see Figure 4.

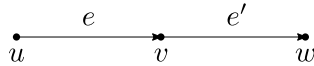


FIGURE 4. $e = (u, v)$ is the predecessor of $e' = (v, w)$

By a *non-backtracking walk* of length k we mean a sequence of directed edges e_0, e_1, \dots, e_k such that $e_i \rightarrow e_{i+1}$, $i = 0, \dots, k-1$. If there exists a non-backtracking walk of length k from e to e' , we write $e \rightarrow_k e'$.

There is a corresponding operator on $\ell^2(E(G))$ called the *non-backtracking operator*. It is usually denoted by $B = B_G$, and is defined by

$$(7) \quad (Bf)(e) := \sum_{e' \rightarrow e} f(e'), \quad f \in \ell^2(E(G)), e \in E(G).$$

For regular graphs its spectrum is closely related to that of the adjacency operator, but the non-backtracking operator often proves to be a more efficient tool, for example in understanding the spectral gap and expansion properties of random regular graphs [8, 15, 28]. In [1], and more recently, in [7, 21], the mixing time and the cutoff phenomenon were examined for non-backtracking random walks.

Here we will need an estimate for the norm of the k -th power of the non-backtracking operator in the special case of $G = T_d$. It is known that the spectral radius of B is $\sqrt{d-1}$ in this case. (See [2] for results on the non-backtracking spectrum for the universal cover of any finite graph.) It immediately follows that

$$\|B^k\| = \left(\sqrt{d-1} + o(1)\right)^k,$$

which implies a bound $\left(\frac{1+o(1)}{\sqrt{d-1}}\right)^k$ for the correlation in Theorem 4.1. We want to prove, however, a more explicit bound in order to get a good quantitative result in Theorem 1.1. To this end we need to more carefully estimate the norm of B^k .

Theorem 4.2. *Let B be the non-backtracking operator of T_d . Then for any positive integer k we have*

$$\|B^k\| \leq (k+1)(\sqrt{d-1})^{k+1}.$$

We postpone the proof until Section 4.5.

4.2. Decomposition of the Koopman representation. In this section we briefly explain how the Γ -action on $L^2(M^S, \mu^S)$ can be decomposed into the sum of quasi-regular representations of Γ . (This can be found in [20, Section 3]. See also [25, Theorem 2.1, Corollary 2.2] and [18, Section 3.2] for the special case when S is the vertex set of a Cayley graph.) This will help us to understand the behavior of the non-backtracking operator on $L^2(M^S, \mu^S)$ in the next section.

A group Γ acts naturally on the left cosets of a subgroup $\Delta \leq \Gamma$. The corresponding Γ -action on $\ell^2(\Gamma/\Delta)$ is called *quasi-regular representation*. If Δ is the trivial subgroup, we get the *regular representation* of Γ on $\ell^2(\Gamma)$. It is easy to see that if Δ is finite (which will always be the case in our setting), then the quasi-regular representation is a subrepresentation of the regular representation, see [20, Lemma 3.3].

Now let M be a measurable space and μ a probability measure on M . Suppose that the Hilbert space $L^2(M, \mu)$ has a countable orthonormal basis: g_0, g_1, g_2, \dots , where g_0 will be assumed to be the constant 1 function. (The same would work for atomic measures μ but with a finite orthonormal basis.) For a countable set S let ν be the product measure μ^S on M^S . First we construct an orthonormal basis for $L^2(M^S, \nu)$. By \mathcal{I} we denote the set of finitely supported $S \rightarrow \{0, 1, 2, \dots\}$ functions. For each $q \in \mathcal{I}$ we define an $M^S \rightarrow \mathbb{R}$ function:

$$W_q(\omega) := \prod_{s \in S} g_{q(s)}(\omega_s) \text{ for any } \omega = (\omega_s)_{s \in S}.$$

Note that this is actually a finite product, since all but finitely many terms are equal to $g_0 \equiv 1$. According to [20, Lemma 3.1] the functions W_q , $q \in \mathcal{I}$ form an orthonormal basis of $L^2(M^S, \nu)$.

Suppose that a countable group Γ acts on S . Recall that (3) defines a Γ -action on M^S , which, in turn, induces a Γ -action on $L^2(M^S, \nu)$:

$$(\gamma \cdot f)(\omega) := f(\gamma^{-1} \cdot \omega), \quad f \in L^2(M^S, \nu), \omega \in M^S, \gamma \in \Gamma.$$

(In representation theory the Γ -action on $L^2(M^S, \nu)$ is called the *Koopman representation*.)

One can define a Γ -action on \mathcal{I} along the same lines:

$$(\gamma \cdot q)(s) := q(\gamma^{-1} \cdot s), \quad q \in \mathcal{I}, s \in S, \gamma \in \Gamma.$$

The Γ -actions on \mathcal{I} and $L^2(M^S, \nu)$ are compatible in the sense that

$$(8) \quad W_{\gamma \cdot q} = \gamma \cdot W_q \text{ for any } \gamma \in \Gamma \text{ and } q \in \mathcal{I}.$$

It follows that the Koopman representation is equivalent to the Γ -action on $\ell^2(\mathcal{I})$. Given any $q \in \mathcal{I}$, the subspace in $\ell^2(\mathcal{I})$ corresponding to the orbit $\Gamma \cdot q = \{\gamma \cdot q : \gamma \in \Gamma\}$ will be invariant under the Γ -action, and the restriction of the action to this subspace is equivalent to the quasi-regular representation on $\ell^2(\Gamma/\Gamma_q)$, where Γ_q denotes the stabilizer of q . It follows that the Γ -action on $\ell^2(\mathcal{I})$ is equivalent to a direct sum of quasi-regular representations [20, Proposition 3.2].

4.3. Generalized non-backtracking operator. We will use the above observations for the case when Γ somehow corresponds to the directed edge set $E(T_d)$ of the d -regular tree. We claim that there exists a subgroup $\Gamma \leq \text{Aut}(T_d)$ that acts *sharply transitively* on the directed edges: for any pair e_1, e_2 of directed edges, Γ has a unique element γ taking e_1 to e_2 . For a simple proof, draw T_d in the plane and consider all graph automorphisms Φ of T_d that preserve the order and the orientation of neighbors (that is, if $v \in V(T_d)$ has neighbors v_1, \dots, v_d in a clockwise order, then $\Phi(v_1), \dots, \Phi(v_d)$ should be the neighbors of $\Phi(v)$ also in a clockwise order). Such automorphisms clearly form a subgroup of $\text{Aut}(T_d)$. It is also easy to see that prescribing the image of a directed edge uniquely determines Φ .

In this section Γ will be a fixed subgroup with the above properties. There is a one-to-one correspondence between Γ and $E(T_d)$: pick a fixed distinguished directed edge \bar{e} , and for any $e \in E(T_d)$ let $\gamma_e \in \Gamma$ be the unique element that takes e to \bar{e} . The next claim describes the group elements corresponding to the predecessors and successors of e .

Claim 4.3. *Let $\bar{e}_1, \dots, \bar{e}_{d-1}$ denote the $d-1$ successors of \bar{e} , that is, $\bar{e} \rightarrow \bar{e}_i$, $i = 1, \dots, d-1$. Then for any directed edge e :*

- (a) $\gamma_{\bar{e}_i} \gamma_e$, $i = 1, \dots, d-1$ correspond to the successors of e ;
- (b) $\gamma_{\bar{e}_i}^{-1} \gamma_e$, $i = 1, \dots, d-1$ correspond to the predecessors of e .

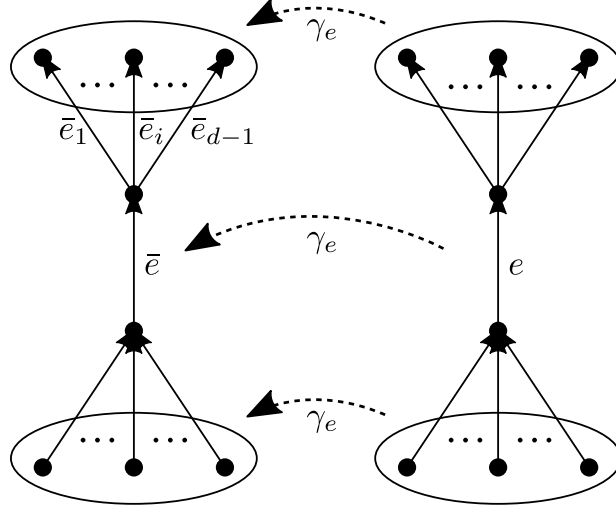
Proof. Since γ_e is a graph automorphism taking e to \bar{e} , it follows that γ_e takes the successors of e to the successors $\bar{e}_1, \dots, \bar{e}_{d-1}$ of \bar{e} , and by definition $\gamma_{\bar{e}_i}$ takes \bar{e}_i to \bar{e} , see Figure 5. This proves (a). Similarly, γ_e takes the predecessors of e to the predecessors of \bar{e} (in some order), and it is easy to see that for any predecessor of \bar{e} there is a unique i for which $\gamma_{\bar{e}_i}^{-1}$ takes that predecessor to \bar{e} , which clearly proves (b). \square

Using the automorphisms $\gamma_{\bar{e}_i}$ we can define some kind of a non-backtracking operator $\mathcal{B} = \mathcal{B}_\Omega$ on $L^2(\Omega, \nu)$ for any Ω with a Γ -action:

$$(9) \quad \mathcal{B}f := \sum_{i=1}^{d-1} \gamma_{\bar{e}_i} \cdot f.$$

If Ω is Γ itself (with the natural action and ν being the counting measure), then $L^2(\Omega, \nu) = \ell^2(\Gamma)$, and using (b) in the above claim we get

$$(\mathcal{B}f)(\gamma_e) = \sum_{i=1}^{d-1} f(\gamma_{\bar{e}_i}^{-1} \gamma_e) = \sum_{e' \rightarrow e} f(\gamma_{e'}),$$

FIGURE 5. Successors and predecessors of e

which means that in this case \mathcal{B} is unitarily equivalent to the ordinary non-backtracking operator B on $\ell^2(E(T_d))$, recall (7).

As for the case when Ω is Γ/Δ for some finite subgroup $\Delta \leq \Gamma$, we saw in the previous section that the Γ -action on $\ell^2(\Gamma/\Delta)$ will be a subrepresentation of the regular representation on $\ell^2(\Gamma)$, therefore \mathcal{B} will be unitarily equivalent to $B|_H$, where B is the non-backtracking operator of T_d and H is some invariant subspace of $\ell^2(E(T_d))$.

Now let Ω be M^S with ν being some product measure μ^S for a countable set S with a Γ -action. The only additional assumption we will need is that the stabilizer of any $s \in S$ is finite.

Let \mathcal{I} be as in the previous section. We have seen that there is an isometry between $L^2(M^S, \nu)$ and $\ell^2(\mathcal{I})$ that preserves the Γ -action. For $0 \equiv q_0 \in \mathcal{I}$, the orbit of q_0 has only one element (hence the stabilizer Γ_{q_0} is the whole group Γ). The corresponding invariant subspace in $L^2(M^S, \nu)$ is the space of constant functions. Let us focus on the restriction \mathcal{B}_0 of the non-backtracking operator \mathcal{B} to the orthogonal complement of the constant functions:

$$L_0^2(M^S, \nu) := \left\{ f \in L^2(M^S, \nu) : \int f \, d\nu = 0 \right\}.$$

For any other $q_0 \neq q \in \mathcal{I}$ the stabilizer Γ_q is finite. It follows from the previous discussion that \mathcal{B}_0 is unitarily equivalent to the direct sum of restrictions of B to various invariant subspaces. Therefore we have the same bound for $\|\mathcal{B}_0^k\|$ as we had for $\|B^k\|$ in Theorem 4.2.

We will also need the following formula for \mathcal{B}^k which is an immediate consequence of part (a) of Claim 4.3 and the definition (9) of \mathcal{B} :

$$(10) \quad \mathcal{B}^k f = \sum_{\bar{e} \rightarrow_k e} \gamma_e \cdot f.$$

4.4. Edge correlations. Now we turn to the proof of Theorem 4.1. Let Γ still denote a subgroup of $\text{Aut}(T_d)$ that acts sharply transitively on the directed edge set $E(T_d)$. Set $M = [0, 1]$ with μ being the Lebesgue measure, and $S = V(T_d)$. (In fact, we can work with arbitrary M, μ , and any S with an $\text{Aut}(T_d)$ -action on it such that the stabilizer of any element intersected with Γ is finite.)

Let $F: M^S \rightarrow \mathbb{R}^{E(T_d)}$ be an $\text{Aut}(T_d)$ -factor. Then for an i.i.d. process Z on M^S , $F(Z) = Y = (Y_e)_{e \in E(T_d)}$ will be a factor of i.i.d. process on $\mathbb{R}^{E(T_d)}$.

Let $\pi_e: \mathbb{R}^{E(T_d)} \rightarrow \mathbb{R}$ denote the coordinate projection corresponding to $e \in E(T_d)$. Recall that \bar{e} is a fixed distinguished directed edge and that $\gamma_e \in \Gamma$ takes e to \bar{e} . As we saw in Section 2, the rule $f := \pi_{\bar{e}} \circ F$ determines F . Straightforward calculation shows that

$$\gamma_e \cdot f = \pi_e \circ F \text{ implying that } Y_e = (\gamma_e \cdot f)(Z).$$

Combining this with (10) we get that

$$(\mathcal{B}^k f)(Z) = \sum_{\bar{e} \rightarrow_k e} Y_e.$$

The assumption in Theorem 4.1 that the variance of Y_e is finite is equivalent to f being in $L^2(M^s, \nu)$. Then $\text{var } Y_e = \|f\|_2^2$ for each $e \in E(T_d)$. Also, since adding a constant does not change the correlation, we might assume that $\mathbb{E}Y_e = 0$, which means that $f \in L_0^2(M^s, \nu)$. Then

$$\langle f, \mathcal{B}^k f \rangle = \mathbb{E}f(Z)(\mathcal{B}^k f)(Z) = \mathbb{E}Y_{\bar{e}} \left(\sum_{\bar{e} \rightarrow_k e} Y_e \right) = \sum_{\bar{e} \rightarrow_k e} \mathbb{E}Y_{\bar{e}} Y_e.$$

Since Y is $\text{Aut}(T_d)$ -invariant, $\mathbb{E}Y_{e_1} Y_{e_2}$ is the same for any edges $e_1 \rightarrow_k e_2$ (that is, any edges e_1, e_2 of distance k and “pointing to the same direction”). Therefore the sum on the right-hand side is actually equal to $(d-1)^k \mathbb{E}Y_{e_1} Y_{e_2} = (d-1)^k \text{cov}(Y_{e_1}, Y_{e_2})$. Using this and the fact that $\|\mathcal{B}_0^k\| \leq \|B^k\|$ (see the previous section), as well as the bound for $\|B^k\|$ in Theorem 4.2 we obtain that

$$\begin{aligned} (d-1)^k |\text{cov}(Y_{e_1}, Y_{e_2})| &= |\langle f, \mathcal{B}^k f \rangle| \leq \|f\|_2 \cdot \|\mathcal{B}^k f\|_2 \leq \|\mathcal{B}_0^k\|_2 \cdot \|f\|_2^2 \\ &\leq (k+1)(\sqrt{d-1})^{k+1} \sqrt{\text{var}(Y_{e_1}) \text{var}(Y_{e_2})}. \end{aligned}$$

Therefore Theorem 4.1 follows for the case when e_1 and e_2 point to the same direction. As for the case when they point away from or towards each other, we need to do the same for the scalar product $\langle \gamma \cdot f, \mathcal{B}^k f \rangle$ and $\langle f, \mathcal{B}^k(\gamma \cdot f) \rangle$, respectively, where $\gamma \in \Gamma$ is the unique element that flips \bar{e} (that is, takes \bar{e} to its inverse). Since $f \mapsto \gamma \cdot f$ is a unitary operator for any fixed γ , we get the same bound for these scalar products as well.

4.5. The norms of the powers of the non-backtracking operator. In this section we give a proof for Theorem 4.2. We will follow the arguments presented in [2, Theorem 4.2] where they bounded the spectral radius of the non-backtracking operator on an arbitrary tree without leaves.

The norm $\|B^k\|$ can be computed as

$$\|B^k\| = \sup_{\|f\|=\|g\|=1} \langle B^k f, g \rangle,$$

where $\|\cdot\|$ and $\langle \cdot, \cdot \rangle$ denote the standard norm and inner product on $\ell^2(E(G))$. Let us expand this inner product using the definition of B .

$$\begin{aligned} \langle B^k f, g \rangle &= \sum_e (B^k f)(e)g(e) = \sum_{e' \rightarrow_k e} f(e')g(e) \leq \sum_{e' \rightarrow_k e} |f(e')g(e)| \\ &\leq \sum_{e' \rightarrow_k e} \frac{1}{2} \left(\frac{1}{\alpha_{e,e'}} f^2(e') + \alpha_{e,e'} g^2(e) \right). \end{aligned}$$

The last bound is based on the inequality of arithmetic and geometric means. Note that there is freedom in choosing the positive constants $\alpha_{e,e'}$ individually for every pair $e' \rightarrow_k e$, which we will discuss later. We may collect the terms $f^2(e')$ and $g^2(e)$ together:

$$(11) \quad \begin{aligned} \langle B^k f, g \rangle &\leq \frac{1}{2} \sum_{e'} \left(\sum_{e' \rightarrow_k e} \frac{1}{\alpha_{e,e'}} \right) f^2(e') + \frac{1}{2} \sum_e \left(\sum_{e' \rightarrow_k e} \alpha_{e,e'} \right) g^2(e) \\ &\leq \frac{1}{2} \sup_{e'} \left(\sum_{e' \rightarrow_k e} \frac{1}{\alpha_{e,e'}} \right) \|f\|^2 + \frac{1}{2} \sup_e \left(\sum_{e' \rightarrow_k e} \alpha_{e,e'} \right) \|g\|^2. \end{aligned}$$

Our goal now is to choose $\alpha_{e,e'}$ in a way that the above suprema are as small as possible. To this end we fix a root o in T_d and define the “spheres” around this root as

$$V_i := \{v \in V \mid d(v, o) = i\}.$$

Similarly, we can partition the edges based on their distance from the root:

$$H_i := \{e = (u, v) \mid u \in V_{i-1}, v \in V_i\} \cup \{e = (u, v) \mid u \in V_i, v \in V_{i-1}\}.$$

This allows us to define the height of an edge as

$$h(e) := i, \text{ if } e \in H_i.$$

Finally, we set

$$\alpha_{e,e'} := \left(\frac{1}{\sqrt{d-1}} \right)^{h(e')-h(e)}.$$

Next we compute what bounds we get for the suprema above.

Claim 4.4. *For any $e' \in E(G)$ we have*

$$\sum_{e' \rightarrow_k e} \frac{1}{\alpha_{e,e'}} < (k+1) \sqrt{d-1}^{k+1}$$

Proof. We need to browse through all the different configurations of e and e' .

If e' is directed away from the root, then all the edges e reached in k steps will be k levels above e' . Indeed, there is no possibility to turn back as these are non-backtracking paths. This way we reach $(d-1)^k$ different edges, and we get

$$\sum_{e' \rightarrow_k e} \frac{1}{\alpha_{e,e'}} = (d-1)^k \left(\frac{1}{\sqrt{d-1}} \right)^k = \sqrt{d-1}^k.$$

Let us now check an edge e' which is pointing towards the root. First we assume that $h(e') > k$. The edges e reached can be either found going all the way down, or taking $l = 0, 1, \dots, k-1$ steps down and then turning up for another $k-l$ steps.

When going simply downwards, there is one such e to reach, with a height decrease of k , thus contributing to the sum by

$$\sqrt{d-1}^k$$

When turning back after l steps, we reach $(d-2)(d-1)^{k-l-1}$ edges, and the height increase is $-l + (k-l-1) = k-1-2l$. The overall contribution for this l is

$$(d-2)(d-1)^{k-l-1} \left(\frac{1}{\sqrt{d-1}} \right)^{k-1-2l} = (d-2) \sqrt{d-1}^{k-1} < \sqrt{d-1}^{k+1}.$$

This bound is valid for all l separately, therefore combining the contributions of all the cases (going all the way down or turning back after $l = 0, 1, \dots, k - 1$ steps) we get

$$\sum_{e' \rightarrow_k e} \frac{1}{\alpha_{e,e'}} < (k+1)\sqrt{d-1}^{k+1}.$$

The only case remaining is when e' is pointing towards the root, but $h(e') \leq k$. It is easy to verify that the above method works again, but some values of l are excluded, and once the intermediate $(d-2)$ factor increases to $(d-1)$. Nevertheless, the same final bound holds.

We checked all the cases for e, e' and confirmed the stated bound for every possibility. \square

Claim 4.5. *For any $e' \in E(G)$ we have*

$$\sum_{e' \rightarrow_k e} \alpha_{e,e'} < (k+1)\sqrt{d-1}^{k+1}$$

Proof. Basically the same proof works as in Claim 4.4, only a small adjustment needs to be made due to the change of orientations. \square

Plugging the bounds from Claim 4.4 and 4.5 into (11) we get

$$\begin{aligned} \langle B^k f, g \rangle &\leq \frac{1}{2}(k+1)\sqrt{d-1}^{k+1} \|f\|^2 + \frac{1}{2}(k+1)\sqrt{d-1}^{k+1} \|g\|^2 \\ &= (k+1)\sqrt{d-1}^{k+1}, \end{aligned}$$

and this is exactly the bound we were aiming for.

REFERENCES

- [1] Noga Alon, Itai Benjamini, Eyal Lubetzky, and Sasha Sodin. Non-backtracking random walks mix faster. *Commun. Contemp. Math.*, 9(4):585–603, 2007.
- [2] Omer Angel, Joel Friedman, and Shlomo Hoory. The non-backtracking spectrum of the universal cover of a graph. *Trans. Amer. Math. Soc.*, 367(6):4287–4318, 2015.
- [3] Ágnes Backhausz and Balázs Szegedy. On large girth regular graphs and random processes on trees. Preprint. arXiv:1406.4420 [math.PR], 2014.
- [4] Ágnes Backhausz, Balázs Szegedy, and Bálint Virág. Ramanujan graphings and correlation decay in local algorithms. *Random Structures Algorithms*, 47(3):424–435, 2015.
- [5] Ágnes Backhausz and Bálint Virág. Spectral measures of factor of i.i.d. processes on vertex-transitive graphs. *To appear in Ann. Inst. Henri Poincaré Probab. Stat.*, arXiv:1505.07412 [math.PR], 2015.
- [6] Karen Ball. Factors of independent and identically distributed processes with non-amenable group actions. *Ergodic Theory Dyn. Syst.*, 25(3):711–730, 2005.
- [7] Anna Ben-Hamou and Justin Salez. Cutoff for non-backtracking random walks on sparse random graphs. Preprint. arXiv:1504.02429 [math.PR], 2015.
- [8] Charles Bordenave. A new proof of Friedman’s second eigenvalue theorem and its extension to random lifts. Preprint. arXiv:1502.04482 [math.CO], 2015.
- [9] Lewis Bowen. The ergodic theory of free group actions: entropy and the f -invariant. *Groups Geom. Dyn.*, 4(3):419–432, 2010.
- [10] Lewis Bowen. Sofic entropy and amenable groups. *Ergodic Theory Dynam. Systems*, 32(2):427–466, 2012.
- [11] Clinton T. Conley, Andrew S. Marks, and Robin Tucker-Drob. Brooks’s theorem for measurable colorings. Preprint. arXiv:1601.03361 [math.CO], 2016.
- [12] Endre Csóka. Independent sets and cuts in large-girth regular graphs. Preprint. arXiv:1602.02747 [math.CO], 2016.
- [13] Endre Csóka, Balázs Gerencsér, Viktor Harangi, and Bálint Virág. Invariant Gaussian processes and independent sets on regular graphs of large girth. *Random Structures Algorithms*, 47(2):284–303, 2015.

- [14] Endre Csóka and Gábor Lippner. Invariant random matchings in cayley graphs. Preprint. arXiv:1211.2374 [math.CO], 2012.
- [15] Joel Friedman. A proof of Alon’s second eigenvalue conjecture and related problems. *Mem. Amer. Math. Soc.*, 195(910):viii+100, 2008.
- [16] Damien Gaboriau and Russell Lyons. A measurable-group-theoretic solution to von Neumann’s problem. *Invent. Math.*, 177(3):533–540, 2009.
- [17] David Gamarnik and Madhu Sudan. Limits of local algorithms over sparse random graphs. *Proceedings of the 5-th Innovations in Theoretical Computer Science conference, ACM Special Interest Group on Algorithms and Computation Theory*, 2014.
- [18] Viktor Harangi and Bálint Virág. Independence ratio and random eigenvectors in transitive graphs. *Ann. Probab.*, 43(5):2810–2840, 2015.
- [19] Carlos Hoppen and Nicholas Wormald. Local algorithms, regular graphs of large girth, and random regular graphs. Preprint. arXiv:1308.0266 [math.CO], 2013.
- [20] A. S. Kechris and T. Tsankov. Amenable actions and almost invariant sets. *Proc. Amer. Math. Soc.*, 136(2):687–697 (electronic), 2008.
- [21] Mark Kempton. Non-backtracking random walks and a weighted Ihara’s theorem. Preprint. arXiv:1603.05553 [math.PR], 2016.
- [22] David Kerr and Hanfeng Li. Soficity, amenability, and dynamical entropy. *Amer. J. Math.*, 135(3):721–761, 2013.
- [23] Gábor Kun. Expanders have a spanning Lipschitz subgraph with large girth. Preprint. arXiv:1303.4982 [math.GR], 2013.
- [24] Russell Lyons. Factors of iid on trees. *Combin. Probab. Comput.*, to appear.
- [25] Russell Lyons and Fedor Nazarov. Perfect matchings as IID factors on non-amenable groups. *European J. Combin.*, 32(7):1115–1125, 2011.
- [26] Donald S Ornstein and Benjamin Weiss. Entropy and isomorphism theorems for actions of amenable groups. *J. Analyse Math.*, 48:1–141, 1987.
- [27] Robin Pemantle. Automorphism invariant measures on trees. *Ann. Probab.*, 20(3):1549–1566, 1992.
- [28] Doron Puder. Expansion of random graphs: new proofs, new results. *Invent. Math.*, 201(3):845–908, 2015.
- [29] Mustazee Rahman. Factor of iid percolation on trees. Preprint. arXiv:1410.3745 [math.PR], 2014.
- [30] Mustazee Rahman and Bálint Virág. Local algorithms for independent sets are half-optimal. Preprint. arXiv:1402.0485 [math.PR], 2014.
- [31] Vladimir Abramovich Rokhlin and Yakov G Sinai. Construction and properties of invariant measurable divisions. *Doklady Akademii Nauk SSSR*, 141(5):1038–1041, 1961.
- [32] Brandon Seward. Weak containment and Rokhlin entropy. Preprint. arXiv:1602.06680 [math.DS], 2016.

EÖTVÖS LORÁND UNIVERSITY, DEPARTMENT OF PROBABILITY AND STATISTICS H-1117 BUDAPEST, PÁZMÁNY PÉTER SÉTÁNY 1/C; AND MTA ALFRÉD RÉNYI INSTITUTE OF MATHEMATICS H-1053 BUDAPEST, REÁLTANODA UTCA 13-15

E-mail address: agnes@cs.elte.hu

MTA ALFRÉD RÉNYI INSTITUTE OF MATHEMATICS H-1053 BUDAPEST, REÁLTANODA UTCA 13-15; AND EÖTVÖS LORÁND UNIVERSITY, DEPARTMENT OF PROBABILITY AND STATISTICS H-1117 BUDAPEST, PÁZMÁNY PÉTER SÉTÁNY 1/C

E-mail address: gerencser.balazs@renyi.mta.hu

MTA ALFRÉD RÉNYI INSTITUTE OF MATHEMATICS H-1053 BUDAPEST, REÁLTANODA UTCA 13-15

E-mail address: harangi@renyi.hu

MTA ALFRÉD RÉNYI INSTITUTE OF MATHEMATICS H-1053 BUDAPEST, REÁLTANODA UTCA 13-15

E-mail address: vizermate@gmail.com